Human Interaction and Emerging Technologies (IHIET-AI 2024), Vol. 120, 2024, 471–481
https://doi.org/10.54941/ahfe1004599

Investigating AI Model Limitations in Recognizing Faces and Bodies in Support of Intelligent Ballroom Dance Education

Lewen Ivy Huang

Mulgrave School, West Vancouver, BC, Canada

ABSTRACT

The advancement in face and body detection algorithms has sparked an interest in using them to assist physical education and sports training, where AI can analyze students' body postures and movements to offer corrective guidance and prevent injuries. However, ballroom dance settings are uniquely different from traditional settings often used for face and body detection. On the one hand, most traditional face and body detection algorithms detect individuals instead of a collaborative dyad. Moreover, specific dancing postures may pose additional challenges for AI algorithms to detect. In order to unlock the power of AI in enhancing real-time feedback for dancers on their movements, postures, and expressions, there needs to be a thorough understanding of the capabilities of AI in analyzing complex dance sequences and identifying subtle nuances in body language. In this work, we examined four widely adopted body and face detection models on their effectiveness in detecting ballroom dancers. We found that these models shared key limitations. First, they are more likely to detect the man than the woman in the dyad, especially when the woman is curved backwards. Second, they often detect the couple as one person, mixing different body sections. Third, errors are frequent when the dancers do not face the camera or when they are wearing specialized costumes. This project offers suggestions to diversify the training datasets of such algorithms to make them suitable for new settings including ballroom dance and offers implications on developing intelligent tutoring systems to support ballroom dance education.

Keywords: Skeleton detection, Ballroom dancing, AI mistakes

INTRODUCTION

The advancement in face and body detection algorithms has sparked an interest in using them to assist in physical education and sports training, where AI can analyze students' body postures and movements to offer corrective guidance and prevent injuries (Semeraro & Turmo, 2022). However, dance settings are uniquely different from traditional settings often used for face and body detection. On the one hand, most traditional face and body detection algorithms detect individuals instead of a collaborative dyad. Moreover, specific dancing postures may pose additional challenges for AI algorithms to detect.

With the unique characteristics of Ballroom dancing, we believe there are several key factors that make the detection of figures more difficult for ballroom dancing compared to alternative contexts, e.g., doing physical workouts. These key factors include:

detecting unique postures of dancers in specific costumes; 2) detecting pairs of individual dancers when they dance in a team all the time.

In order to unlock the power of AI in enhancing real-time feedback for dancers in support of intelligent ballroom dance education, there needs to be a thorough understanding of the capabilities of AI in analyzing complex dance sequences and identifying subtle nuances in body language. In this work, we aim to examine a list of widely adopted body and face detection models on their effectiveness in detecting dancers in a specific dance genre – ballroom dance. We gathered a dataset of 30 images of dyads of ballroom dancers and examined four widely adopted body and face detection models on their effectiveness in detecting ballroom dancers. We found that these models shared key limitations. First, they are more likely to detect the man than the woman in the dyad, especially when the woman is curved backwards. Second, they often detect the couple as one person, mixing different body sections. Third, errors are frequent when the dancers do not face the camera or when they are wearing specialized costumes. This project offers suggestions to diversify the training datasets of such algorithms to make them suitable for new settings including ballroom dance.

RELATED WORK

In this section, we review prior work on using AI and digital technologies to enhance dancing practice. One line of work explored providing targeted dance video tutorials. As an example, Semeraro & Turmo (2022) used videos to give learners personalized feedback such as directional cues, body highlights, and visual metaphors.

There has been a plethora of computer vision models developed to capture dance motions and poses, which are used as the basis for AI to give dancers feedback (Zhai, 2021; Pang & Niu, 2023; Jin, 2023; Alarcon, 2018; Yan & He, 2022; Kanjee, 2022). Deb et al. (2018) developed a platform that uses deep learning algorithms to extract a 2D skeleton model to improve dance learning for children, the dance positions are displayed to the user, and the user must replicate the same pose. The system captures the user's image using a camera and compares it to the original dance position to give users feedback. In terms of teaching beginners to dance, researchers have developed gamified platforms to give personalized feedback to players. For example, Tang, Chan & Leung (2011) proposed a dance game that could identify the player's dance motions in real time and have a virtual partner recognize and respond to the player's movement. In another project, Kang et al. (2023) proposed a 2D online dance learning system DancingInside that helps beginners to learn dance by delivering timely and enough feedback based on Teacher-AI collaboration.

Researchers have also developed algorithms to detect multiple dancers when they are dancing together. For example, Emily Carson (CBC News, 2020; Big think, 2020) developed technology to capture the motions of multiple dancers with 94% accuracy. Other research showed that when there are multiple dancers on the scene, the position could be inaccurate when the dancers obscure each other (Tian & Yang, 2022a; Tian & Yang, 2022b).

Recent work has also explored using AI to generate dance moves. For example, Liu & Ko (2022) developed systems that can provide corresponding dance action candidates and combine them into a new dance action, increasing the smoothness and rationality of the long-time dance sequence. In another work (Emerging Technology, 2017), researchers developed a game that allowed players to dance to the music required by the game, which enabled gamers to create and share their own dances.

The application of AI in dancing is not limited to pose estimation and motion detection, more work has looked into head gesture and faction expression detection. For example, The Dance Coach (Romano, Schneider & Drachsler, 2019) uses a Kinect V2 to detect the learner's faction expressions and body movements in order to recognize the dancing moves and offer feedback. Another project (Nussipbekov, Amirgaliyev & Hahn, 2014) developed the Hidden Markov Chain and included headwear as a new skeleton joint. This enhances the accuracy of head gesture identification in a dancer's whole-body gesture recognition.

CONTEXT

This research is conducted in the context of competitive Ballroom dancing. In competitive Ballroom dancing (Marion, 2008), dancers are judged based on a diverse set of criteria and several key factors, including musicality, posture and gesture, body alignment, partnership, techniques and overall performance (NDCA, 2023; Emeraldball, 2023). Musicality is the first thing dancers are evaluated on, dancers are required to notice the counts and dance on counts based on the syllabus book or open-level rules. Judges will observe how well the dancers match their steps and choreography to the rhythm, phrasing, and dynamics of the music. Secondly, judges consider the overall presentation and styling of the dancers. Posture and gesture are an expression of your dance representing an identity of yourself, having your own personal attitude and symbol is important. Additionally, if a dancer is able to maintain a posture that is higher than their level, it will make them stand out. A good posture with good costume choice and poise of attention to detail enhances the overall visual presentation and the impact of the performance. In ballroom dancing, the partnership is fundamental. For all levels, judges pay attention to how well the dancing couple works together as a team. They consider factors such as communication, lead and follow techniques, and synchronization with each other's movements. For higher levels, as the dancers are more experienced, the judges will look at factors including responsiveness and connection. Since ballroom syllabus dancers follow a route, it might cause multiple dancers to be in the same area. Responsiveness suggests that when such an incident happens, the man in the couple should lead and the follower should understand the intentions of the leader. Connection refers to the force that a couple applies to each other, e.g., using their partner to make their speed faster, movements larger, and have better balance for stretching. Lastly, judges and experienced dancers assess the dancers' technical proficiency, including elements such as footwork, frame, and control. Attention to detail and execution of specific dance steps, figures, and movements are crucial in demonstrating solid technique.

Based on these unique characteristics of Ballroom dancing, we believe there are several key factors that make the detection of figures and pose more difficult for ballroom dancing compared to alternative contexts. These key factors include: 1) detecting unique postures of dancers in specific costumes; 2) detecting pairs of individual dancers when they dance in a team all the time.

METHOD

Definition of Postures

Our work starts with defining the ballroom posture from the syllables book for Waltz's basic posture that specifies how two dancers should connect to each other (Marion, 2008). Usually, Latin dances are formed by two people, a man and a lady as a dance couple. The man is referred to as the leader and the lady is referred to as the follower. The two dancers use their hands to apply some force toward each other, to help themselves to reach their postures and make their frame more stable. Instead of pulling the partner towards yourself and losing balance, applying some force to the partner on the hands' frame will be the correct way. Latin generally focuses more on speed and body and muscles' extrusion, while Ballroom focuses more on postures. Both dancers have their own posture which they should maintain while dancing the route.

The general posture of the couple is the same, however, depending on the movements and the shaping in ballroom dancing, sometimes we allow postures to change. In the general posture, the man is straight, making sure of the direction and safety of the dancers as the lead. The lady is the follower which is the extension of the man's movement with the posture of the right side of the man and woman's hip touching together, both of the dancers with soft knees bend for moving quickly, and the woman has her leg beside the man's right leg as close but not limiting both of the dancers' movement. Her body is wrapped around the man's right side until her rib cage and her top above is like a high jump, as shown in the right image of Figure 1. She is not bending backwards, but instead looking upwards. The lady is responsible for holding her upper body and maintaining balance in the posture.

There is a level scale for the lady's posture, measured by the distance between the man's face and the lady's. Level zero means that both of the partners are facing each other, and both are straight. Level ten means that the man is straight with the lady bent to 90 degrees (as shown in the right image of Figure 1).

Figure 1: Two examples show the level scale used to assess a couple's postures in ballroom dancing. The left image shows a level 0 where both dancers are straight, and the right image shows a level 10 where the lady is bent to 90 degrees.

Using AI to Detect Dancer Postures

The goal of this research is to examine to what extent AI algorithms can accurately detect dancers' figures and skeletons for Ballroom dancing, in order to develop downstream applications to give automated feedback to dancers.

In this research, we examined 4 state-of-the-art models for face and skeleton detection, including two skeleton detection and post- estimation models, including the Face Plus Plus skeleton detection model (https://www.faceplusplus.com/) and the MediaPipe pose landmark detection model (https://mediapipe-studio.webapps.google.com/studio/demo/pose_landmarker), and face detection models, including MediaPipe face detection mode (https://mediapipe-studio.webapps.google.com/studio/demo/face_detector), and the DETR object detection model released through HuggingFace (https://huggingface.co/facebook/detr- resnet-50).

Dataset Preparation

We prepared a dataset of 31 images composed of good and bad examples of ballroom dancing. We randomly selected 31 ballroom waltz dancers online. For good examples, we have the criteria of having the posture of level 10, having the correct posture and gesture, and having an elegant overall image. The bad examples may demonstrate errors such as the elbows being fully straight, especially when it feels like the leader is pulling the follower's right hand up into the air, the faces facing the wrong direction, and the two dancers' core connections are separate or facing directly towards each other. Some of the bad examples picked in our dataset show extremely wrong postures from the general posture.

The image dataset on ballroom dancing we prepared is representative of a typical ballroom dancing posture by showing a variety of dancers as examples. The typical ballroom dancing posture does not have a fixed image, as many people all have their highlights and feelings, but they all share the same base. Dancers form techniques, frame, and add their style of gesture and atmosphere around them, so that even though all of them are doing the same figure, it looks different in details and overall images.

Experiments

With the image dataset and the 4 models, I tried experimenting with using each of the 4 models to detect skeletons or faces in each of the images. I created a table for each model and labelled the AI output as success or failure. Specifically, I labelled the success based on whether the dancers in the image are successfully detected by the AI.

Data Analysis

For the skeleton and post-detection models (Face plus plus and Mediapipe), we measure their success based on whether they can successfully label the dancer's joints and body parts correctly by having the skeleton line on the dancer's body parts, having points on the dancer's joints, and its successful prediction of the dancer's arms and legs. If all of the above is detected successfully, we count it as successful, even though the AI sometimes only detects one of the dancers. Otherwise, we take notes based on the parts AI detects as wrong and make hypotheses of the reasons that lead to the mistakes.

For the face detection models (DETR from Hugging Face and Mediapipe), we measure whether it can successfully separate both dancers by using a rectangle to frame the dancers apart. If only one of the dancers is framed out, it's considered unsuccessful, and sometimes the AI even detects background noise and objects successfully. For mediapipe only, we consider the AI detection to not only successfully frame both dancers out but also label them clearly as people. Otherwise, it is unsuccessful. We will analyze the AI detection output in the note section to see what factors cause the AI to get the opposite detection output as what we expected.

FINDINGS

We then analyze the results to investigate the errors and we summarize findings on when the models work or do not work. We find that in ballroom dancing, the legs of both dancers are often concealed in the images. Specifically, the women's legs are typically covered by their dresses, whereas sometimes one of the men's legs is visible, making it challenging for the AI to identify the dancers' legs accurately. However, the visibility of the man's leg tends to make him more detectable compared to the woman. Additionally, the man assumes a straight posture while the woman bends, which further contributes to the man being more easily detectable by the AI.

In all four models, we observed some common factors that cause AI to detect the images unsuccessfully. This includes 1) when the dancers are too close to each other, AI detects them as the same person. 2) The posture and position of the dancers and the angle of the camera taking the photo, for example, AI is more accurate when people are straight towards the camera. 3) Some environmental elements or photography properties may be associated with mistakes in detection as well. For example, when there are people or chairs in the background, when the photo's colour scheme is too simple, and when the photo is blurry, there are more likely to be errors.

Detecting two people as one person. We found that the most common mistake made by both the Face plus plus model and the Mediapipe model was that the two dancers were detected as the same person. As shown in the Figure to the left, the skeleton detection model perceives both dancers as one entity when the lady's clothing partially obscures the central area of the man's body. In such a case, the AI has difficulty distinguishing between the body parts of the two persons. In the Mediapipe examples (bottom two images), where the woman dancer is facing the camera, the model often mixes the female dancer's upper body with the male dancer's lower body.

Sideway camera angle leads to errors. When the dancers are facing sideways to the camera, the AI is more likely to make mistakes.

Men are detected more often than women. We found that men dancers were more often detected than the women dancers. For example, when the male dancer is facing the camera or when the male dancer maintains a straight posture, AI is more likely to detect them correctly, as shown in the Figure above. In ballroom dancing, men are more likely to maintain a straight posture and face the camera, in our experiment, we found that men dancers are more likely to be picked up by an AI model.

Predictions of the hidden lower body are often based on a standing posture. We also observed cases where the model was using the dancer's upper body including hand and arm directions to predict their hidden lower body. Such predictions are assuming the person is standing straight, but putting that in a dancing context, the predictions are often wrong.

Half-body view conducive to errors. We observed that when the photo only presented a half-body view of the dancers in contrast to full-body views, the detection was often incorrect.

Unfamiliar environmental and photography elements lead to errors. Other factors may also have contributed to the errors. For example, bird eye view makes it hard for AI to detect correctly. The AI may mistake a dancer's hair as a body section. This might be because, in the training dataset of the models, there are not sufficient training examples that are either bird eye view or present diverse hairstyles as we have observed in the dataset. Additionally, we observed that when the photo is black-and-white (has fewer colours), or when there are many people in the background, the model makes more errors. For example, the model may pick up people in the background more than the dancers in the front as shown in the images.

AI can detect specific body sections correctly but not the whole body. Even when the model is not able to detect both dancers accurately, we observed that in some cases, the model was able to capture a specific body section of one dancer accurately. For example, in the images shown on the left, although the model detects both dancers as the same person, the upper body detection is often accurate, especially the arms and shoulders, and the face detection of the dancer who is facing the camera is often correct.

DISCUSSION

Our experiments revealed several areas of mistakes in using AI models to detect human body skeletons and faces in a ballroom dancing setting. We further analyzed the potential sources of these errors. One important finding was that the AI models seemed to have a tendency of picking up on men dancers over women dancers. This may also be because men dancers often maintain a straight posture which is more prevalent in the training examples, in contrast to women dancers who often were bent towards the back. Moreover, women dancers' lower bodies were often covered by their dresses, making them invisible by the AI. Given that ballroom dancing is more prevalent in the Latin culture, where dancers commonly engage in skin tanning or use darker foundations, we recognized that these practices could pose challenges for the AI in accurately classifying dancers.

We found common mistakes made by the AI models we experimented with. It fails to accurately predict hidden parts of the dancing that are covered by the dancers' dresses. Additionally, it struggles with detecting and differentiating the body parts of both dancers. Moreover, the model lacks accuracy due to factors such as the colour scheme, half-body images, the woman's non- straight posture, and the varying angles of the camera. To address these shortcomings, we encourage researchers to improve the accuracy of AI detection systems on basic dancing postures. We have identified specific features and characteristics that AI detection can be enhanced to mitigate the mistakes and biases. These include: 1) accurately separate the two dancers working in a team; 2) consider the camera angles of the images; 3) consider the unique postures for dancers and the unique costumes dancers may wear; 4) be more adaptive to half-body views.

We imagine that better-performing machine learning models for ballroom dance detection can be trained with curated datasets. For example, images labelled on the skeletons for both the lead and the follower dancer instead of just one of them. By diversifying the training datasets to include images of different colour schemes, camera angles (e.g., having dancers face sideways, having partners' back facing the camera, only showing the upper body of the dancers), and dancers in different competitive levels (the lady's bending scale will range from 0 to 10), the performance of AI in detecting the two dancers will be significantly improved. Such models will empower downstream applications, e.g., to give personalized and automated feedback to ballroom dancers. Better algorithms on detecting and differentiating dancers could serve as the basis for developing intelligent tutoring systems in support of ballroom dancing.

REFERENCES

Alarcon, N. (2018). AI Can Transform Anyone Into a Professional Dancer. https://developer.nvidia.com/blog/ai-can-transform-anyone-into-a-professional-dancer/

Big think. (2020). New AI can identify you by your dancing "fingerprint". https://bigthink.com/high-culture/a-i-dancing-fingerprint/.

CBC News. (2020). Forget face recognition — an AI can tell who you are by how you dance Social Sharing. https://www.cbc.ca/radio/quirks/feb-1-understanding-the-coronavirus-cyborg-jellyfish-judging-cat-pain-and-more-1.5446302/forget-face-recognition-an-ai-can-tell-who-you-are-by-how-you-dance-1.5446316.

Dance Movement Recognition Based on Modified GMM-Based Motion Target Detection Algorithm. Security and Communication Networks, 2022.

Deb, S., Sharan, A., Chaturvedi, S., Arun, A., & Gupta, A. (2018). Interactive dance lessons through human body pose estimation and skeletal topographies matching. International Journal of Computational Intelligence & IoT, 2(4).

Emeraldball. (2023). Rules & Policies. https://emeraldball.com/rules-policies/.

Emerging Technology. (2017). Machine-Learning Algorithm Watches Dance Dance Revolution, Then Creates Dances of Its Own. https://www.technologyreview.com/2017/03/28/106144/machine-learning-algorithm-watches-dance-dance-revolution-then-creates-dances-of-its-own/.

Jin, G. (2023). Application of image recognition based on neural network algorithm in multimedia dance teaching. Soft Computing, 1–9.

Kang, J., Kang, C., Yoon, J., Ji, H., Li, T., Moon, H.,... & Han, J. (2023). Dancing on the inside: A qualitative study on online dance learning with teacher-AI cooperation. Education and Information Technologies, 1–31.

Kanjee, R. (2022). Top 9 Pose Estimation Models of 2022. https://medium.com/augmented-startups/top-9-pose-estimation-models-of-2022-70d00b11db43.

Marion, J. S. (2008). Ballroom: Culture and costume in competitive dance. Berg.

NDCA. (2023). NDCA Rules & Regulations Febuary 2023 Edition. https://www.ndca.org/pdf/2023%20February%20-%20Compiled%20Rule%20Book%20Master%20v2.pdf.

Nussipbekov, A. K., Amirgaliyev, E. N., & Hahn, M. (2014, April). Kazakh traditional dance gesture recognition. In Journal of Physics: Conference Series (Vol. 495, No. 1, p. 012036). IOP Publishing.

Pang, Y., & Niu, Y. (2023). Dance Video Motion Recognition Based on Computer Vision and Image Processing. Applied Artificial Intelligence, 37(1), 2226962.

Romano, G., Schneider, J., & Drachsler, H. (2019). Dancing Salsa with Machines—Filling the Gap of Dancing Learning Solutions. Sensors, 19(17), 3661.

Semeraro, A., & Turmo Vidal, L. (2022, April). Visualizing instructions for physical training: Exploring visual cues to support movement learning from instructional videos. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1–16).

Tang, J. K., Chan, J. C., & Leung, H. (2011, February). Interactive dancing game with real-time recognition of continuous dance moves from 3D human motion capture. In Proceedings of the 5th international conference on ubiquitous information management and communication (pp. 1–9).

The use of deep learning technology in dance movement generation. Frontiers in Neurorobotics, 16, 911469.

Yan, M., & He, Z. (2022). Dance Action Recognition Model Using Deep Learning Network in Streaming Media Environment. Journal of Environmental and Public Health, 2022.

Zhai, X. (2021). Dance movement recognition based on feature expression and attribute mining. Complexity, 2021, 1–12.